



Ye, J., Gillespie, K. M., & Rodriguez, S. (2018). Unravelling the roles of susceptibility loci for autoimmune diseases in the post-GWAS era. *Genes*, 9(8), [377]. <https://doi.org/10.3390/genes9080377>

Publisher's PDF, also known as Version of record

License (if available):  
CC BY

Link to published version (if available):  
[10.3390/genes9080377](https://doi.org/10.3390/genes9080377)

[Link to publication record in Explore Bristol Research](#)  
PDF-document

This is the final published version of the article (version of record). It first appeared online via MDPI at <http://www.mdpi.com/2073-4425/9/8/377> . Please refer to any applicable terms of use of the publisher.

## University of Bristol - Explore Bristol Research

### General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:  
<http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>

Review

# Unravelling the Roles of Susceptibility Loci for Autoimmune Diseases in the Post-GWAS Era

Jody Ye <sup>1,\*</sup>, Kathleen M. Gillespie <sup>1</sup> and Santiago Rodriguez <sup>2</sup> 

<sup>1</sup> Diabetes and Metabolism, Translational Health Sciences, University of Bristol, Level 2 Learning and Research Building, Southmead Hospital, Bristol BS10 5NB, UK; k.m.gillespie@bristol.ac.uk

<sup>2</sup> MRC Integrative Epidemiology Unit, Population Health Sciences, University of Bristol, Bristol BS8 2BN, UK; santi.rodriguez@bristol.ac.uk

\* Correspondence: jody.yi.ye@bristol.ac.uk; Tel.: +44-(0)-117-414-8041

Received: 9 May 2018; Accepted: 23 July 2018; Published: 27 July 2018



**Abstract:** Although genome-wide association studies (GWAS) have identified several hundred loci associated with autoimmune diseases, their mechanistic insights are still poorly understood. The human genome is more complex than single nucleotide polymorphisms (SNPs) that are interrogated by GWAS arrays. Apart from SNPs, it also comprises genetic variations such as insertions-deletions, copy number variations, and somatic mosaicism. Although previous studies suggest that common copy number variations do not play a major role in autoimmune disease risk, it is possible that certain rare genetic variations with large effect sizes are relevant to autoimmunity. In addition, other layers of regulations such as gene-gene interactions, epigenetic-determinants, gene and environmental interactions also contribute to the heritability of autoimmune diseases. This review focuses on discussing why studying these elements may allow us to gain a more comprehensive understanding of the aetiology of complex autoimmune traits.

**Keywords:** complex loci; autoimmune diseases

## 1. Introduction

Autoimmune disease is a major human health burden, affecting 5% to 8% of the world's population. To date, more than 80 autoimmune diseases have been described [1]. For some of the more common conditions, such as multiple sclerosis, Crohn's disease, rheumatoid arthritis, and type 1 diabetes, increases in incidence and prevalence have been observed in westernised societies over the last several decades [2–5]. Both genetic and environmental factors are thought to play roles in the initiations and progression of autoimmunity. Environmental factors such as viral infection [6], nutrition [7], gut dysbiosis [8] and in utero environment [9] have all been postulated to play a role but have been difficult to confirm due to variations among individuals, populations and geographical areas [10–15]. In contrast, common genetic components of autoimmune diseases have been better characterised.

In the 1970s, studies of twins and first-degree relatives identified the most important genetic risk attributed by the human leukocyte antigen (HLA) region [16–20]. In the subsequent decades, candidate gene studies led to the discovery of several other predisposing loci for autoimmune diseases. For example, variable number tandem repeat (VNTR) class I alleles upstream of the insulin gene were associated with type 1 diabetes [21–23]; *PTPN22* was found to be associated with rheumatoid arthritis, type 1 diabetes, systemic lupus erythematosus, and Grave's disease [24]; *CTLA4* was associated with a number of autoimmune conditions including autoimmune thyroid disease, type 1 diabetes, coeliac disease [25]; and *IL2RA* was associated with type 1 diabetes, rheumatoid arthritis, and multiple sclerosis [2,26,27]. The subsequent genome-wide association studies (GWAS) comparing affected cases

with unrelated healthy individuals substantially increased the power of gene discovery analyses for autoimmune diseases, which led to over 300 loci being identified [28]. GWAS improved our knowledge of disease risk, but the specific design of GWAS does not allow consideration of other elements potentially involved in disease susceptibility. This is because GWAS are based on common genetic variants, specifically single nucleotide polymorphisms (SNPs). While the aetiologies of ‘complex diseases’ are attributed by genetic and environmental components, as well as the interplay between the both, genetic factors themselves are comprised of SNPs and genetic variations at many other levels. In addition, phenomena such as gene-gene interactions (epistasis), and genetic-epigenetic interactions increase the complexity of the genetic basis of human common diseases and traits. Recent studies suggest that many causal variants trigger autoimmune responses in a cell-type specific and/or cell-state specific manner. In this review, we discuss the roles of complex loci at genetic and epigenetic levels in the aetiology of autoimmune diseases.

## 2. The Genome-Wide Association Studies Era

Genome-wide Association Studies were facilitated by advancing technology to conduct high-throughput SNP analyses in very large case-control populations. For autoimmune diseases, by identifying susceptible genetic variants, one can (a) help understand the underlying biological pathways to inform the design of novel immune therapies; and (b) predict the risk of individuals developing autoimmunity. One of the strengths of GWAS is its study design. The powerful hypothesis-free, association mapping design of GWAS has enabled the identification of hundreds of candidate genes strongly associated with human traits. Briefly, a GWAS is performed first by genotyping a set of SNPs (often over 500 k SNPs) using commercial microarrays. Subsequently, genotypes at the genome-wide level are imputed according to the haplotype structures provided by reference panels. For example, based on the early HapMap phase 1–3 reference panel, the Wellcome Trust Case Control Consortium published a number of GWAS for common autoimmune diseases such as rheumatoid arthritis and type 1 diabetes [29]. The 2.5 million SNPs from the HapMap reference panel only included the most common variants, with minor allele frequencies greater than 5% of the study population [30]. The subsequent 1000 Genomes Project interrogated genotypes of nearly 40 million variants, including 1.4 million insertion/deletion variants [31], which dramatically improved GWAS coverage, allowing the identification of several novel autoimmune disease loci [32]. One of the applications of GWAS results is the development of genetic risk score (GRS). Genetic risk score does not aim to detect individual SNPs, but instead is an aggregate of genetic risk across the genome. It is calculated by combining the effect sizes of multiple SNPs, weighted by the strength of each SNP [33]. Although GRS is not powerful enough to make clinical diagnosis at individual level [34], it has been shown to be useful for patient stratification and risk prediction [35–38]. For example, in type 1 diabetes, it was suggested that targeting the top 18% individuals of the general population with the highest ranked GRS would capture 80% of future cases [39]. Given that the prevalence of type 1 diabetes is 0.4% in the general population, this will decrease the number of individuals that will need to be treated to prevent one type 1 diabetes case from 250 to 50 [39].

## 3. Missing Heritability in Genome-Wide Association Studies

Narrow sense heritability ( $h^2$ ) is a term that defines the proportion of phenotypic variance that is contributed by additive genetic variance [40]. It has been shown that GWAS tend to explain a smaller fraction of  $h^2$  compared to those estimated from classical methods based on population data, such as from the analysis of offspring and parental phenotypes, siblings or monozygotic and dizygotic twins [41]. This gap is known as missing heritability [41]. Even in highly heritable conditions such as type 1 diabetes and juvenile idiopathic arthritis ( $h^2$  up to 90%), SNP based  $h^2$  still falls short of those reported from population estimates [41]. Missing heritability was originally thought to be partly contributed by rare variants of the genome. To examine this hypothesis, Hunt et al. first performed dense exon sequencing of 20 GWAS loci that exhibit shared susceptibilities to six common autoimmune

diseases to identify putative rare variants and then conducted single-variant association analyses among the 20 loci; their findings suggested that rare variants have a negligible impact on autoimmune disease risk [42]. Thus, the components of missing heritability remain largely unknown and may involve complex loci, gene-gene interactions and epigenetic regulations [43,44].

#### 4. The Role of Complex Loci at Genetic Level

Complex loci include many types of genetic variations, such as insertion-deletion polymorphisms (indels), VNTRs (including microsatellites and minisatellites), copy number variants (CNVs, with each DNA fragment that is ~1 kb or larger) [45], long interspersed nuclear elements [46], short interspersed nuclear elements, Alu repeats, somatic mosaicism, and cytogenetic abnormalities (including insertions, duplications, translocations, inversions). There is evidence demonstrating functional effects on human diseases caused by complex genetic variations of the genome [46,47]. One particular example is the VNTR region upstream of the *INS* promoter, as it was one of the very first loci to be associated with type 1 diabetes [48]. Of the three subclasses of *INS* VNTR, class I homozygous alleles were predisposing (odds ratio (OR) = 2.68) and class III alleles were dominantly protective (OR = 0.37) [21]. The predisposing class I variants were associated with lower levels of *INS* mRNA expression in the thymus, implying impaired tolerance induction of T-cells during thymic education [49]. Another example is the VNTR of the *C4* (complement component 4), which has been associated with systemic lupus erythematosus. Systemic lupus erythematosus is an autoimmune condition where deficiency in the clearance of apoptotic cells and immune complexes is involved as part of the disease aetiology [50]. The protein product of *C4* (*C4A*) plays a pivotal role in the activation of the classical and the lectin complement pathways that lead to cytolysis or neutralisation of invading microbes, clearance of immune complexes and apoptotic materials [51]. Low copy number of *C4* VNTR was strongly associated with increased risk of systemic lupus erythematosus and was coupled with lower levels of plasma *C4A*, whereas high copy number of *C4* was protective and was associated with higher levels of plasma *C4A* [51].

Due to the complexity of structural DNA elements, technical challenges exist to study them directly in large-scale populations. A simple way to profile their associations in common diseases is through tagged SNPs; for example, a given SNP in linkage disequilibrium with another genetic variant, such as a VNTR or CNV, could capture the genetic variations and therefore be used as a proxy. However, it is possible that not all of the structural elements are well-tagged by SNP arrays used to conduct GWAS, and hence, they are not directly tested in GWAS for their associations with autoimmune diseases. Sequencing studies showed that CNVs account for a major proportion of human genetic polymorphism [52,53]. To assess whether CNVs can explain missing heritability of common diseases, the Wellcome Trust Case Control Consortium employed custom arrays that interrogate approximately 50% of all common CNVs (>5% frequency) and performed association analyses for eight diseases, including bipolar disorder, breast cancer, coronary artery disease, Crohn's disease, hypertension, rheumatoid arthritis, type 1 diabetes and type 2 diabetes; over 3400 polymorphic CNVs were surveyed [54]. Although the study identified three associated CNV loci—*HLA* for Crohn's disease, rheumatoid arthritis and type 1 diabetes; *IRGM* for Crohn's disease; and *TSPAN8* for type 2 diabetes, all of which were tagged by SNPs from previous GWAS [54]. However, a limitation of this study was that CNVs that were poorly assigned into distinct classes were eliminated from the analyses. A subsequent study by Zanda et al. further tested 3410 previously unclustered and untagged CNVs on their associations with type 1 diabetes using a family study design [55]. The authors observed no convincing associations between untagged CNVs and type 1 diabetes. Therefore, data from both studies led to the conclusion that common CNVs do not explain a major proportion of missing heritability. If CNVs are involved in susceptibility to common autoimmune diseases, one would imagine that they must be rare variants with large effect sizes. In fact, only 22% of rare CNVs (<1% frequency) were tagged by at least one SNP, indicating that they would be interesting targets for investigation [54]. Cooper et al. investigated whether rare CNVs could have a significant contribution to type 1 diabetes risk [56]. In their study, 6808 type 1 diabetes and 9954 control individuals were

compared; de novo rare CNV deletions (rDELs) were found more frequently in type 1 diabetes patients. In addition, rare CNVs were divided into small (0–20 kb), medium (20–400 kb), and large (>400 kb) groups; there was a greater proportion of large rDELs in type 1 diabetes patients (OR = 1.57). More interestingly, very long CNVs were found in eleven type 1 diabetes cases and only one control (rDELs > 3 Mb, OR = 15.60). These rDELs are enriched in genes that regulate immune function. Because of the rarity of CNVs, the study had insufficient power and coverage. Nevertheless, these data still imply that rare CNVs are potentially involved in the risk for type 1 diabetes, but it was estimated that over 30,000 cases and equivalent number of controls are required to detect rare CNVs with large effects (ORs approximately 8) [56].

Another phenomenon that potentially contributes to missing heritability of autoimmune disease is somatic mosaicism. Unlike being passed from parents or developed ‘de novo’ in sperm or egg, this type of mutation occurs post-zygotically after the single cell stage, resulting in two genetically distinct cell populations within an individual [57]. Somatic mosaicism can be found in monozygotic twins; particularly, it was demonstrated in twins with neurological disorders [58], implying that it may play a role in disease pathogenesis. Somatic mosaicism has also been described in a number of autoimmune diseases. For example, de novo CNVs were previously found in the affected twins with primary biliary cirrhosis of monozygotic twin pairs [59]. Another example is autoimmune lymphoproliferative syndrome (ALPS), which is mainly characterised by the proliferation of CD4-/CD8- double negative lymphocytes [60]. These double negative T cells are originated from activated single positive T cells that had heterozygous somatic mutations of *Fas*, resulting in their resistance to apoptosis [61]. More rarely, somatic mosaicism has been reported to associate with a combination of autoimmune diseases. For example, a *KRAS* mutation was detected in T and B lymphocytes but not in natural killer cells of a patient with multiple autoimmunity, including thrombocytopenia, recurrent Henoch-Schonlein purpura and intestinal Behçet disease [62]. Further analyses are required to determine whether rare genetic variations such as rDELs and somatic mosaicism have significant additional value to SNPs in missing heritability of autoimmune diseases or if their effects are negligible. Other possible factors explaining missing heritability may include inaccurate heritability estimates, study design limitations of GWAS (such as sample size, type of genetic variants considered), non-additive genetic effects (epistasis, gene and environment interaction), epigenetics and disease heterogeneity [43,63,64].

Genome-wide Association Studies are set out to detect the most associated signals within a given genomic region. Interestingly, some regions such as the HLA locus demonstrate multiple association signals with autoimmune diseases. To better define independent association signals, conditional stepwise regression is often used. This was designed to ask the question whether a single variant could best explain a trait and conditional on this single variant, whether any other variants explain additional trait variance [65]. The international multiple histocompatibility complex and autoimmunity genetics network performed conditional regression analyses on seven autoimmune disorders in the HLA region [66]. For example, in systemic lupus erythematosus, the top associated signal was rs1269852, a SNP located between *TNXB* and *CREBL1* in chromosome 6, in strong linkage disequilibrium with *HLA-DRB1\*0301*. Conditioning on the top variant identified multiple secondary association signals such as *HLA-DRB1\*1501*, class I (between *RNF39* and *TRIM31*), class II (*HLA-DQB1-DQA2*), and class III (*NOTCH4*) HLA alleles. For Crohn’s disease, conditioning on the top signal *HLA-DRB1\*1101* identified additional independent associations such as rs382259 (located near the *NOTCH4* region), the class I region (between *HLA-B* and *MICA*) and the *DQA-DQB1-DQA2-DQB2* region. For multiple sclerosis, conditional analysis revealed that apart from the top association *DRB1\*1501* (tagged by rs3135391), *HLA-B\*4402* and *HLA-C\*0501* also appeared to have independent associations. Together, these data demonstrate that the HLA region exhibits complex and multi-locus effect for autoimmune associations.

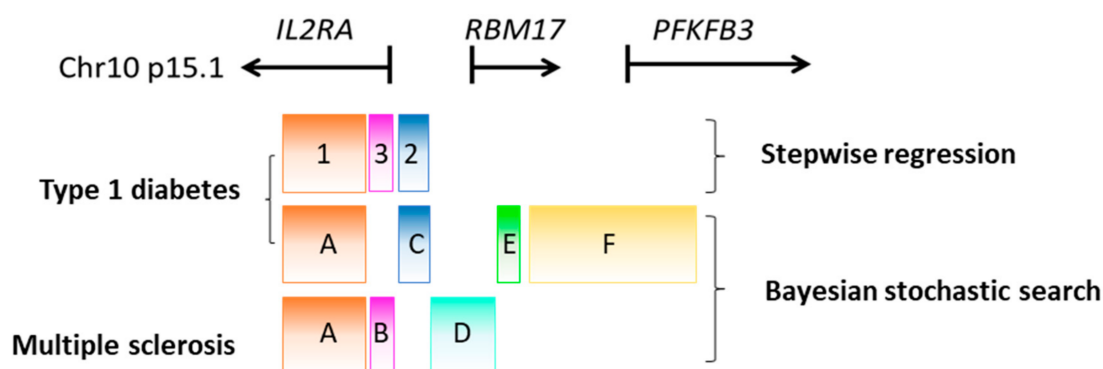
## 5. Identifying Candidate Causal Variants Using Fine Mapping

Genome-wide Association Studies signals highlight regions of associations, but the lead variants may not be causal to disease phenotype. Identifying causal variants in regions associated with autoimmune diseases is



a challenging task, often due to high linkage disequilibrium structure and multiple disease-causing variants in relatively close proximity. Therefore, various fine-mapping strategies have been developed and we chose *IL2RA* as an example to illustrate how fine-mapping strategies have helped to define candidate causal variants for several autoimmune diseases. The *IL2RA* (CD25) is located on chromosome 10p15.1; it encodes a subunit (IL-2R $\alpha$ ) of the receptor for the pro-inflammatory cytokine IL-2, which has been associated with a number of autoimmune diseases including multiple sclerosis, rheumatoid arthritis, autoimmune thyroid disease, and type 1 diabetes [2,27,67]. Upon stimulation of IL-2, IL-2 receptor signals to maintain the suppressive functions of CD4+FOXP3+ regulatory T cells and facilitate effector and memory T cell differentiation [68]. The *IL2RA* region was initially found to be associated with type 1 diabetes using a multi-locus genetic association test in 2005 [69]. Using conditional logistic regression as mentioned previously, Lowe et al. fine-mapped the region covering *IL2RA* and its neighbouring gene *RBM17*, which led to the identification of two loci independently associated with type 1 diabetes [70], each comprising a number of indistinguishable SNPs. Group 1 is located in intron1 of *IL2RA* (marked by rs12722495, previously marked by rs41295061) and Group 2 is located at the intergenic region between the 5' of *IL2RA* and *RBM17* (marked by rs11594656) [70]. Maier et al. later discovered an independent Group 3 signal in intron 1 of *IL2RA*, tagged by SNP rs2104286 [71]. Apart from type 1 diabetes, Group 3 is also associated with multiple sclerosis. Functionally, the protective variant of the Group 1 SNP rs12722495 induces higher expression levels of CD25 on the surface of CD4+ memory T cells, potentially causing increased T cell activation in response to IL-2 stimulation. The protective variant of Group 3 SNP rs2104286 was counterintuitively associated with a lower percentage of CD25+ cells in CD4+ naïve fraction, suggesting a reduced likelihood of T cell activation [72].

The recent development of fine-mapping methods such as Bayesian stochastic search proved to be more efficient in detecting multiple independent association signals. Unlike conditional logistic regression, Bayesian stochastic search tests the question which sets of SNPs can best jointly explain type 1 diabetes association. Wallace et al. applied this approach to re-analyse the *IL2RA* region in 2015. They showed that, instead of three, there are four groups of SNPs independently associated with type 1 diabetes [65], with Group A SNPs located in the intron 1 of *IL2RA* (that is equivalent to Group 1), Group C SNPs located in the intergenic region between *IL2RA* and *RBM17* (equivalent to Group 2), which replaced the previous Group 3 SNP, Group E SNPs located at the 5' of *RBM17*, and Group F SNPs resided in the 5' of *RBM17* to intron 2 of *PFKFB3* (Figure 1) [65]. For multiple sclerosis, the risk could be explained either by Group A and Group D SNPs (tagged by rs56382813) jointly or by Group B SNP (rs2104286) alone.



**Figure 1.** Schematic representation of multiple independent associations at the *IL2RA*-*RBM17*-*PFKFB3* region discovered using stepwise regression and Bayesian stochastic search that are associated with both type 1 diabetes and multiple sclerosis. Group 1–3 single nucleotide polymorphisms that contribute to type 1 diabetes were originally identified using stepwise regression. The later-developed Bayesian stochastic search, however, identified four groups of SNPs (A, C, E, F) that jointly explain type 1 diabetes risk at the *IL2RA* locus; whereas for multiple sclerosis, the risk at this locus was explained by Group A and D jointly, or by Group B SNPs alone. Figure was adapted and modified from [65].

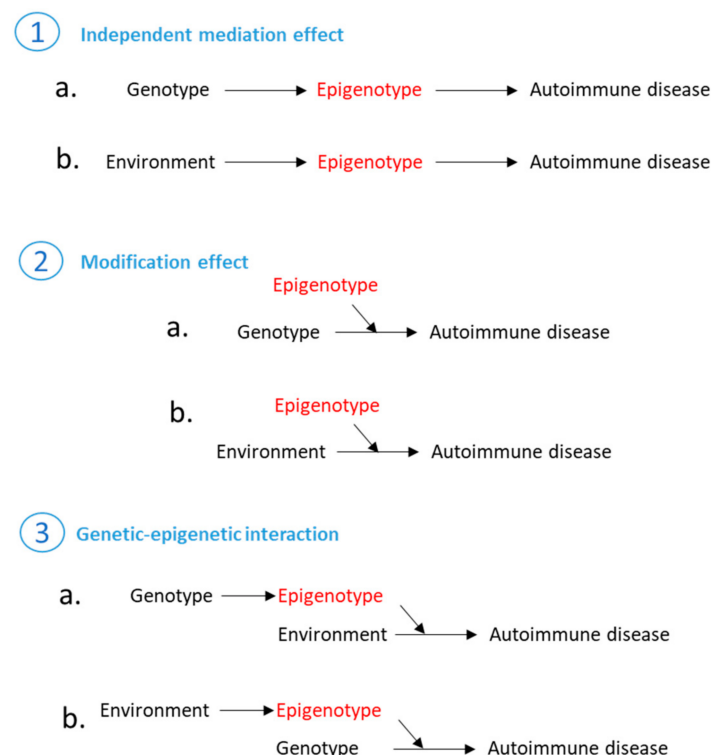
## 6. Epistasis in Autoimmune Diseases

Genes may not function in isolation. Gene–gene interactions (epistases) are major contributors to autoimmune disease risk; a classical example is the HLA class II haplotypes. For example, *HLA-DRB1\*1501-DQB1\*0602* is the most susceptible haplotype for multiple sclerosis [73] and *HLA-DRB1\*0301-DQB1\*0201* is the most susceptible haplotype for type 1 diabetes [74]. A summary of HLA class II haplotypes in autoimmune diseases is reviewed here [75]. Class II and class I interactions also play a role in autoimmune disease risk. For instance, in type 1 diabetes patients, a combination of *HLA-A\*24*, *DQA1\*03*, and *DR9* has been associated with accelerated beta cell loss [76]. The interaction between *HLA-DR3/DR4* and class I *A\*03* allele demonstrated significant protective effect of clinical progression to type 1 diabetes, whereas *HLA-DR3/DR4* and class I *B\*39* interaction contributed significantly to the progression from multiple islet autoantibody to type 1 diabetes [77]. HLA also interact with non-HLA genes. For example, an increased risk of systemic lupus erythematosus (OR = 1.19) was observed when *CTLA4* (cytotoxic T lymphocyte antigen 4), a negative regulator of T cell response interacts with SNPs rs3131379 and rs1270942 located in the HLA class III region [78]. Another similar example was reported in multiple sclerosis. A driver of increased risk in multiple sclerosis is the soluble form of interleukin-7 receptor (IL-7R, encoded by *IL7R*). Exon 6 of *IL7R* interacts with many proteins, one of them is encoded by the *DDX39B* that is located in the HLA region. Galarza-Munoz et al. recently identified that a SNP rs2523506 within the *DDX39B* region reduces *DDX39B* expression. More importantly, the authors showed a significant increased risk in multiple sclerosis when rs2523506 interacts with a risk variant of *IL7R* (rs6897932). Carriers of risk alleles of both SNPs increased soluble IL-7R expression, thereby increasing the disease risk [79]. Although epistasis at the HLA region is readily detectable because HLA genes are the main susceptibility genes for autoimmune diseases, interactions between HLA and non-HLA genes, as well as between two non-HLA genes are difficult to observe due to the limited power at the current GWAS sample size. This is mainly explained by two reasons. Firstly, there is limited evidence that epistasis between the HLA and non-HLA genes, as well as between two non-HLA genes comprise a large fraction of total genetic variation in autoimmune diseases. Secondly, in the additive effect model, the loss of information between genotyped SNPs and causal variants is proportional to linkage disequilibrium ( $r^2$ ), but in epistasis, the loss of information between the two is proportional to  $r^4$  [80]. Therefore, a much larger sample size is necessary to detect epistasis than to detect the main effect in GWAS [81].

## 7. Epigenetic Regulation

Epigenetic mechanism is another layer of regulation that influences a gene function. Although the Greek prefix ‘epi’ indicates an effect that is acting ‘on’ the genome, epigenotype can influence autoimmune susceptibility in several ways, as illustrated in Figure 2. As shown in the first scenario in Figure 2, epigenetic mechanisms can independently mediate genetic and environmental risk, which subsequently lead to autoimmune disease. Perhaps the most well-studied epigenetic mechanisms involve DNA methylation, histone modification, and long non-coding RNAs. DNA methylation occurs at the cytosine-phosphate-guanine (CpG) residues; it has been shown that approximately 20% of DNA methylation variance is explained by additive genetic variance [82]. Single nucleotide polymorphisms that are in close proximity to the CpG site (cis, <1 Mb centred on the SNP) as well as SNPs that are far away from the CpG site (trans, >1 Mb centred on the SNP) can both influence DNA methylation levels [82,83]. These SNPs are known as methylation quantitative trait loci (mQTLs). Many early studies attempted to find mQTL associated DNA methylation changes by performing epigenome-wide association analyses [84–86]. However, one needs to carefully interpret these results, because it is difficult to discriminate causality (that DNA methylation is causally influencing a trait) from reverse causation (that DNA methylation is a consequence of changes in gene expression), linkage disequilibrium confounding (that causal variants and mQTLs are simply in linkage disequilibrium) or from horizontal pleiotropy (that mQTLs alter DNA methylation and autoimmune diseases via different mechanisms) [87]. Recent developments in statistical strategies such as causal inference

test [88], Mendelian randomisation (MR) [89,90], and genetic co-localisation fine mapping [87] have made such discriminations possible.



**Figure 2.** Schematic representation of three modes of actions where epigenetic regulations can take place to contribute to autoimmune disease risk. Figure was adapted and modified from [91].

Liu and colleagues applied a causal inference test on a cohort of 354 rheumatoid arthritis patients and 337 controls to investigate whether genetic risk of rheumatoid arthritis is mediated by DNA methylation. They identified a large number of differentially methylated CpG sites within the HLA region (535 SNP-CpG pairs) that potentially exhibit mediatory effect. A similar observation was made at a non-HLA locus, within the Glutathione S-Transferase Alpha 2 (*GSTA2*) gene [92]. Other studies using the same approach have found that DNA methylation potentially mediate the expression of a number of genes in human islets, including *HLA-DQB1*, the main predisposing gene to type 1 diabetes [93]. In our recent work, we applied Mendelian Randomisation together with genetic co-localisation fine-mapping to study whether DNA methylation mediates the genetic risk of type 1 diabetes. We identified a number of loci, including *CTSH*, *PTPN2* and *AFF3*, where DNA methylation is potentially on the causal pathway to type 1 diabetes [94]. Richardson et al. further extended this statistical framework to systematically investigate the functional roles of DNA methylation in hundreds of traits including many other autoimmune diseases [95]. DNA methylation appeared to increase autoimmune disease risk in a number of susceptible loci, where inflammatory bowel disease had the greatest number of DNA methylation mediated loci, followed by rheumatoid arthritis and Crohn's disease [96].

DNA methylation may also play a role in the female predominance observed in autoimmune diseases [97]. Part of the reason for the female predominance was thought to be due to a high frequency of skewed X-chromosome inactivation. In females, one copy of the X-chromosome was silenced by DNA methylation whereas the expressed copy of X-chromosome was unmethylated. The choice of which copy of X-chromosome to be silenced is random, which results in two distinct cell populations, where 50% paternal X-chromosome genes are expressed in one cell population and 50% maternal



X-chromosome genes are expressed in another. As a result, the dosage of proteins transcribed from X-chromosomes is approximately equal between males (XY) and females (XX). However, not all women have a 50:50 ratio of cells with one or the other X-chromosomes active. A deviation from equal inactivation of each parental allele is known as skewing; for example, some alleles could be inactivated in 70–80% cells and in extreme cases, in 90–95% cells. This may lead to an altered dosage of proteins to be translated in some females. In peripheral blood mononuclear cells and multiple other tissues, skewed X-chromosome inactivation has been observed more frequently in female patients with autoimmune diseases [98,99]. It is considered that lowered expression of X-linked self-tolerance genes in the thymus may lead to lack of exposure of self-antigens, subsequently leading to the escape of autoreactive T cells.

In the second scenario, an epigenotype is thought to modulate a genetic/environmental risk factor. A typical example of this is the regulation by long non-coding RNAs (lncRNAs). Long non-coding RNAs are RNAs that exceed 200 nucleotides in length and they are broadly classified into five subclasses: stand-alone lncRNAs (or large intergenic non-coding RNAs, lincRNAs), natural antisense transcripts (NATs), pseudogenes-derived lncRNAs [100], long-intronic lncRNAs, and promoter/enhancer-associated lncRNAs [101]. lncRNAs have been shown to involve in anti-viral responses [102], T cell differentiation [103], and NFκB signalling [104]. Mirza et al. overlapped known lncRNAs with susceptible variants for inflammatory bowel disease and type 1 diabetes based on their physical locations in the genome. They identified over 2000 inflammatory bowel disease-associated SNPs physically located within 468 lncRNAs and over 1000 type 1 diabetes SNPs within 247 lncRNAs; many of them potentially disrupt the secondary structure of lncRNAs [105]. The authors therefore hypothesised that some of the autoimmune disease associated SNPs can alter the expression and function of lncRNAs, which subsequently influence disease related genes. This hypothesis was systematically investigated using statistical approaches by Kumar et al. where the authors found that disease associated expression quantitative trait loci (eQTLs) affect 112 out of 2140 lncRNAs in whole blood [106]. Hrdlickova et al. mapped lncRNAs expressed in seven immune cell types (granulocytes, monocytes, NK cells, B cells, memory T cells and naïve CD8+ cells) to susceptible loci in nine autoimmune diseases [107]. They found that the proportion of lncRNAs expressed in autoimmune disease loci were significantly higher than the proportion of lncRNAs expressed genome-wide; additionally, the expression levels of lncRNAs in autoimmune disease loci were higher than that detected at the genome-wide level [107]. More interestingly, lncRNAs overlapping the disease regions tend to be tissue-specific. For example, inflammatory bowel disease associated lncRNAs are preferentially expressed in NK cells, juvenile idiopathic arthritis associated lncRNAs are enriched in memory and CD8+ T cells [107].

The third scenario elucidates the most complicated situation, which is genetic-epigenetic interaction. This type of interaction has been seen in many occasions involving enhancer regulations. Enhancers are defined as cis-acting DNA sequences that can increase the transcription of genes. Ninety eight percentage of enhancers are located in the non-coding regions of the genome, either upstream or downstream of genes, or in introns. Enhancers can be identified using high-throughput sequencing targeting specific markers such as H3 acetylated at lysine 27 (H3K27ac) and H3 monomethylated at K4 (H3K4me1), which are chemical modifications of the histone proteins that wrap around DNA. Their activities can also be specific to a tissue or a particular cell type, a time-point in life, or a unique physiological state [108]. A recent study overlapped causal variants with histone marks in 21 common autoimmune diseases showed that causal variants of a disease trait are enriched in enhancers specific to disease-affecting tissues. For example, causal variants of Alzheimer's disease are enriched in enhancers in the brain; causal variants of type 1 diabetes are enriched in enhancers in lymphocytes as well as in pancreatic islets [109]. In addition, a disproportionate number of enhancers respond to ex vivo stimulation, reflected by increased H3K27Ac (marks active promoter and enhancer) signals and non-coding RNA transcription upon immune cell activation [109].

A specific example of genetic-epigenetic interaction was previously described in Grave's disease. One hypothesis for the initiation of Grave's disease is viral infection. Infection can lead to the

recognition of auto-antigens via molecular mimicry, which further causes bystander activation of auto-reactive T cells and global pro-inflammatory cytokine production [110–112]. To mimic the consequence of viral infection, Stefan et al. treated human thyroid cells with pro-inflammatory cytokines [113]. They observed that thyroid cells exerted significant changes of H3K4me1 signatures at the intron1 of *TSHR*, which harbours a previously predicted causal variant, rs12101261, to Grave's disease. After pro-inflammatory cytokine treatment, rs12101261 was able to interact with histone deacetylase and a transcription repressor PLZF, resulting in reduced *TSHR* expression and breakdown of central tolerance [113].

An immediate problem of studying genetic-epigenetic interaction is that only 10–20% causal variants were predicted to disrupt transcription factor binding motifs at the enhancer sites, 80–90% causal variants function by modifying the non-classical regulatory sequence [109,114]. In addition, the nature of stimulus-dependent enhancer interactions makes it challenging to robustly study them in un-stimulated cells. The advances in Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-Cas9 technologies enabled their screening and characterisation more readily in un-stimulated cells. CRISPR activation (CRISPRa) utilises guide RNAs conjugated with a strong transcriptional activator (i.e., VP64) to induce the expression of endogenously weakly expressed genes [115]. Simeonov et al. recently used this approach to scan enhancers surrounding the *CD69* and *IL2RA* [116]. They identified a CRISPR responsive enhancer at the intronic region of *IL2RA*, and confirmed the ability of the candidate causal variant rs61839660 (group A SNP) to disrupt this enhancer activity in a stimulus dependent manner. Using mouse models, they subsequently showed that upon T cell stimulation, this enhancer controls CD4 naïve T cell polarisation, as CD4+ naïve T cells in the enhancer deletion mouse strain tend to favour a pro-inflammatory Th17 cell differentiation rather than T regulatory cell differentiation [116]. Interestingly, enhancer disruption of *IL2RA* delayed its expression, which was eventually recovered three days after T cell stimulation [116], implying that the induction of autoimmunity could happen in a transient and tissue specific manner.

## 8. Conclusions

With GWAS studies, we have gained significant knowledge to broadly define autoimmune disease-associated regions genome-wide. Although GWAS have their inherent limitations, it was a big step forward, considering that the concept of autoimmunity was initially proposed during the 1940s [117] and the HLA associations were only first described in the 1970s. The challenge for the next decade is to precisely characterise the functions of disease risk loci. With fast-growing profiling of the genomic and epigenomic regulatory elements, as well as cutting edge bioinformatic and molecular genetic tools available, we are in a unique position to study complex loci. Only by understanding the complex and dynamic nature of autoimmune disease mechanisms, can we deliver truly translational research that impact future clinical care.

**Funding:** J.Y. was funded by a Diabetes Research & Wellness Foundation non-clinical fellowship N-C/2016/Ye.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Campbell, A.W. Autoimmunity and the Gut. *Autoimmune Dis.* **2014**. [[CrossRef](#)] [[PubMed](#)]
2. Barrett, J.C.; Clayton, D.G.; Concannon, P.; Akolkar, B.; Cooper, J.D.; Erlich, H.A.; Julier, C.; Morahan, G.; Nerup, J.; Nierras, C.; et al. Rich, and consortium type 1 diabetes genetics. genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. *Nat. Genet.* **2009**, *41*, 703–707. [[CrossRef](#)] [[PubMed](#)]
3. Kappelman, M.D.; Moore, K.R.; Allen, J.K.; Cook, S.F. Recent trends in the prevalence of Crohn's disease and ulcerative colitis in a commercially insured US population. *Dig. Dis. Sci.* **2013**, *58*, 519–525. [[CrossRef](#)] [[PubMed](#)]

4. Benito-Leon, J. Are the prevalence and incidence of multiple sclerosis changing? *Neuroepidemiology* **2011**, *36*, 148–149. [[CrossRef](#)] [[PubMed](#)]
5. Myasoedova, E.; Crowson, C.S.; Kremers, H.M.; Therneau, T.M.; Gabriel, S.E. Is the incidence of rheumatoid arthritis rising? Results from Olmsted County, Minnesota, 1955–2007. *Arthritis Rheum.* **2010**, *62*, 1576–1582. [[CrossRef](#)] [[PubMed](#)]
6. Ercolini, A.M.; Miller, S.D. The role of infections in autoimmune disease. *Clin. Exp. Immunol.* **2009**, *155*, 1–15. [[CrossRef](#)] [[PubMed](#)]
7. Manzel, A.; Muller, D.N.; Hafler, D.A.; Erdman, S.E.; Linker, R.A.; Kleinewietfeld, M. Role of western diet in inflammatory autoimmune diseases. *Curr. Allergy Asthma Rep.* **2014**, *14*. [[CrossRef](#)] [[PubMed](#)]
8. Vieira, S.M.; Pagovich, O.E.; Kriegel, M.A. Diet, microbiota and autoimmune diseases. *Lupus* **2014**, *23*, 518–526. [[CrossRef](#)] [[PubMed](#)]
9. Fronczak, C.M.; Baron, A.E.; Chase, H.P.; Ross, C.; Brady, H.L.; Hoffman, M.; Eisenbarth, G.S.; Rewers, M.; Norris, J.M. In utero dietary exposures and risk of islet autoimmunity in children. *Diabetes Care* **2003**, *26*, 3237–3242. [[CrossRef](#)] [[PubMed](#)]
10. Schmidt, C.W. Questions persist: Environmental factors in autoimmune disease. *Environ. Health Perspect.* **2011**, *119*, A249–A253. [[CrossRef](#)] [[PubMed](#)]
11. Selmi, C.; Tsuneyama, K. Nutrition, geoepidemiology, and autoimmunity. *Autoimmun. Rev.* **2010**, *9*, A267–A270. [[CrossRef](#)] [[PubMed](#)]
12. Stojanovich, L. Stress and Autoimmunity. *Autoimmun. Rev.* **2010**, *9*, A271–A276. [[CrossRef](#)] [[PubMed](#)]
13. Knip, M.; Honkanen, J. Modulation of type 1 diabetes risk by the intestinal microbiome. *Curr. Diabetes Rep.* **2017**, *17*. [[CrossRef](#)] [[PubMed](#)]
14. Watad, A.; Azrielant, S.; Bragazzi, N.L.; Sharif, K.; David, P.; Katz, I.; Aljadeff, G.; Quaresma, M.; Tanay, G.; Adawi, M.; et al. Seasonality and autoimmune diseases: The contribution of the four seasons to the mosaic of autoimmunity. *J. Autoimmun.* **2017**, *82*, 13–30. [[CrossRef](#)] [[PubMed](#)]
15. Hu, Y.; Wong, F.S.; Wen, L. Antibiotics, gut microbiota, environment in early life and type 1 diabetes. *Pharmacol. Res.* **2017**, *119*, 219–226. [[CrossRef](#)] [[PubMed](#)]
16. Singal, D.P.; Blajchman, M.A. Histocompatibility (HL-A) antigens, lymphocytotoxic antibodies and tissue antibodies in patients with diabetes mellitus. *Diabetes* **1973**, *22*, 429–432. [[CrossRef](#)] [[PubMed](#)]
17. Cudworth, A.G.; Woodrow, J.C. Letter: HL-A antigens and diabetes mellitus. *Lancet* **1974**, *2*. [[CrossRef](#)]
18. Stastny, P. Mixed lymphocyte cultures in rheumatoid arthritis. *J. Clin. Investig.* **1976**, *57*, 1148–1157. [[CrossRef](#)] [[PubMed](#)]
19. Moser, K.L.; Kelly, J.A.; Lessard, C.J.; Harley, J.B. Recent insights into the genetic basis of systemic lupus erythematosus. *Genes Immun.* **2009**, *10*, 373–379. [[CrossRef](#)] [[PubMed](#)]
20. Schlosstein, L.; Terasaki, P.I.; Bluestone, R.; Pearson, C.M. High association of an HL-A antigen, W27, with ankylosing spondylitis. *N. Engl. J. Med.* **1973**, *288*, 704–706. [[CrossRef](#)] [[PubMed](#)]
21. Bennett, S.T.; Wilson, A.J.; Cucca, F.; Nerup, J.; Pociot, F.; McKinney, P.A.; Barnett, A.H.; Bain, S.C.; Todd, J.A. IDDM2-VNTR-encoded susceptibility to type 1 diabetes: Dominant protection and parental transmission of alleles of the insulin gene-linked minisatellite locus. *J. Autoimmun.* **1996**, *9*, 415–421. [[CrossRef](#)] [[PubMed](#)]
22. Lucassen, A.M.; Julier, C.; Beressi, J.P.; Boitard, C.; Froguel, P.; Lathrop, M.; Bell, J.I. Susceptibility to insulin dependent diabetes mellitus maps to a 4.1 kb segment of DNA spanning the insulin gene and associated VNTR. *Nat. Genet.* **1993**, *4*, 305–310. [[CrossRef](#)] [[PubMed](#)]
23. Bell, G.I.; Horita, S.; Karam, J.H. A polymorphic locus near the human insulin gene is associated with insulin-dependent diabetes mellitus. *Diabetes* **1984**, *33*, 176–183. [[CrossRef](#)] [[PubMed](#)]
24. Rawlings, D.J.; Dai, X.; Buckner, J.H. The role of PTPN22 risk variant in the development of autoimmunity: Finding common ground between mouse and human. *J. Immunol.* **2015**, *194*, 2977–2984. [[CrossRef](#)] [[PubMed](#)]
25. Gough, S.C.; Walker, L.S.; Sansom, D.M. CTLA4 gene polymorphism and autoimmunity. *Immunol. Rev.* **2005**, *204*, 102–115. [[CrossRef](#)] [[PubMed](#)]
26. Martin, J.E.; Carmona, F.D.; Broen, J.C.; Simeon, C.P.; Vonk, M.C.; Carreira, P.; Rios-Fernandez, R.; Espinosa, G.; Vicente-Rabaneda, E.; Tolosa, C.; et al. The autoimmune disease-associated *IL2RA* locus is involved in the clinical manifestations of systemic sclerosis. *Genes Immun.* **2012**, *13*, 191–196. [[CrossRef](#)] [[PubMed](#)]

27. Stahl, E.A.; Raychaudhuri, S.; Remmers, E.F.; Xie, G.; Eyre, S.; Thomson, B.P.; Li, Y.; Kurreeman, F.A.; Zhernakova, A.; Hinks, A.; et al. Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. *Nat. Genet.* **2010**, *42*, 508–514. [[CrossRef](#)] [[PubMed](#)]
28. Gutierrez-Arcelus, M.; Rich, S.S.; Raychaudhuri, S. Autoimmune diseases—Connecting risk alleles with molecular traits of the immune system. *Nat. Rev. Genet.* **2016**, *17*, 160–174. [[CrossRef](#)] [[PubMed](#)]
29. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3000 shared controls. *Nature* **2007**, *447*, 661–678.
30. Long, T.; Hicks, M.; Yu, H.C.; Biggs, W.H.; Kirkness, E.F.; Menni, C.; Zierler, J.; Small, K.S.; Mangino, M.; Messier, H.; et al. Whole-genome sequencing identifies common-to-rare variants associated with human blood metabolites. *Nat. Genet.* **2017**, *49*, 568–578. [[CrossRef](#)] [[PubMed](#)]
31. de Vries, P.S.; Sabater-Lleal, M.; Chasman, D.I.; Trompet, S.; Ahluwalia, T.S.; Teumer, A.; Kleber, M.E.; Chen, M.H.; Wang, J.J.; Attia, J.R.; et al. Comparison of hapmap and 1000 genomes reference panels in a large-scale genome-wide association study. *PLoS ONE* **2017**, *12*. [[CrossRef](#)] [[PubMed](#)]
32. Li, Y.R.; Li, J.; Zhao, S.D.; Bradfield, J.P.; Mentch, F.D.; Maggadottir, S.M.; Hou, C.; Abrams, D.J.; Chang, D.; Gao, F.; et al. Meta-analysis of shared genetic architecture across ten pediatric autoimmune diseases. *Nat. Med.* **2015**, *21*, 1018–1027. [[CrossRef](#)] [[PubMed](#)]
33. Marees, A.T.; de Kluiver, H.; Stringer, S.; Vorspan, F.; Curis, E.; Marie-Claire, C.; Derks, E.M. A Tutorial on conducting genome-wide association studies: Quality control and statistical analysis. *Int. J. Methods Psychiatr. Res.* **2018**, *27*. [[CrossRef](#)] [[PubMed](#)]
34. Wray, N.R.; Yang, J.; Hayes, B.J.; Price, A.L.; Goddard, M.E.; Visscher, P.M. Pitfalls of predicting complex traits from SNPs. *Nat. Rev. Genet.* **2013**, *14*, 507–515. [[CrossRef](#)] [[PubMed](#)]
35. Sorosina, M.; Esposito, F.; Guaschino, C.; Clarelli, F.; Barizzzone, N.; Osiceanu, A.M.; Brambilla, P.; Mascia, E.; Cavalla, P.; Gallo, P.; et al. Inverse correlation of genetic risk score with age at onset in bout-onset and progressive-onset multiple sclerosis. *Mult. Scler.* **2015**, *21*, 1463–1467. [[CrossRef](#)] [[PubMed](#)]
36. Patel, K.A.; Oram, R.A.; Flanagan, S.E.; de Franco, E.; Colclough, K.; Shepherd, M.; Ellard, S.; Weedon, M.N.; Hattersley, A.T. Type 1 diabetes genetic risk score: A novel tool to discriminate monogenic and type 1 diabetes. *Diabetes* **2016**, *65*, 2094–2099. [[CrossRef](#)] [[PubMed](#)]
37. Oram, R.A.; Patel, K.; Hill, A.; Shields, B.; McDonald, T.J.; Jones, A.; Hattersley, A.T.; Weedon, M.N. A type 1 diabetes genetic risk score can aid discrimination between type 1 and type 2 diabetes in young adults. *Diabetes Care* **2016**, *39*, 337–344. [[CrossRef](#)] [[PubMed](#)]
38. Johnson, M.B.; Patel, K.A.; de Franco, E.; Houghton, J.A.L.; McDonald, T.J.; Ellard, S.; Flanagan, S.E.; Hattersley, A.T. A type 1 diabetes genetic risk score can discriminate monogenic autoimmunity with diabetes from early-onset clustering of polygenic autoimmunity with diabetes. *Diabetologia* **2018**, *61*, 862–869. [[CrossRef](#)] [[PubMed](#)]
39. Polychronakos, C.; Li, Q. Understanding type 1 diabetes through genetics: Advances and prospects. *Nat. Rev. Genet.* **2011**, *12*, 781–792. [[CrossRef](#)] [[PubMed](#)]
40. Zaitlen, N.; Kraft, A.P. Heritability in the genome-wide association era. *Hum. Genet.* **2012**, *131*, 1655–1664. [[CrossRef](#)] [[PubMed](#)]
41. Li, Y.R.; Zhao, S.D.; Li, J.; Bradfield, J.P.; Mohebnasab, M.; Steel, L.; Kobie, J.; Abrams, D.J.; Mentch, F.D.; Glessner, J.T.; et al. Genetic sharing and heritability of paediatric age of onset autoimmune diseases. *Nat. Commun.* **2015**, *6*. [[CrossRef](#)] [[PubMed](#)]
42. Hunt, K.A.; Mistry, V.; Bockett, N.A.; Ahmad, T.; Ban, M.; Barker, J.N.; Barrett, J.C.; Blackburn, H.; Brand, O.; Burren, O.; et al. Negligible impact of rare autoimmune-locus coding-region variants on missing heritability. *Nature* **2013**, *498*, 232–235. [[CrossRef](#)] [[PubMed](#)]
43. Zuk, O.; Hechter, E.; Sunyaev, S.R.; Lander, E.S. The mystery of missing heritability: Genetic interactions create phantom heritability. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 1193–1198. [[CrossRef](#)] [[PubMed](#)]
44. Kilpinen, H.; Dermitzakis, E.T. Genetic and epigenetic contribution to complex traits. *Hum. Mol. Genet.* **2012**, *21*, R24–R28. [[CrossRef](#)] [[PubMed](#)]
45. Freeman, J.L.; Perry, G.H.; Feuk, L.; Redon, R.; McCarroll, S.A.; Altshuler, D.M.; Aburatani, H.; Jones, K.W.; Tyler-Smith, C.; Hurles, M.E.; et al. Copy number variation: New insights in genome diversity. *Genom. Res.* **2006**, *16*, 949–961. [[CrossRef](#)] [[PubMed](#)]

46. Ricano-Ponce, I.; Zhernakova, D.V.; Deelen, P.; Luo, O.; Li, X.; Isaacs, A.; Karjalainen, J.; di Tommaso, J.; Borek, Z.A.; Zorro, M.M.; et al. Refined mapping of autoimmune disease associated genetic variants with gene expression suggests an important role for non-coding rnas. *J. Autoimmun.* **2016**, *68*, 62–74. [[CrossRef](#)] [[PubMed](#)]
47. Strachan, T.; Read, A.P.; Strachan, T. *Human Molecular Genetics*, 4th ed.; Garland Science: New York, NY, USA, 2011.
48. Stead, J.D.; Buard, J.; Todd, J.A.; Jeffreys, A.J. Influence of allele lineage on the role of the insulin minisatellite in susceptibility to type 1 diabetes. *Hum. Mol. Genet.* **2000**, *9*, 2929–2935. [[CrossRef](#)] [[PubMed](#)]
49. Pugliese, A.; Zeller, M.; Fernandez, A., Jr.; Zalcberg, L.J.; Bartlett, R.J.; Ricordi, C.; Pietropaolo, M.; Eisenbarth, G.S.; Bennett, S.T.; Patel, D.D. The insulin gene is transcribed in the human thymus and transcription levels correlated with allelic variation at the *INS* VNTR-*IDDM2* susceptibility locus for type 1 diabetes. *Nat. Genet.* **1997**, *15*, 293–297. [[CrossRef](#)] [[PubMed](#)]
50. Mok, C.C.; Lau, C.S. Pathogenesis of systemic lupus erythematosus. *J. Clin. Pathol.* **2003**, *56*, 481–490. [[CrossRef](#)] [[PubMed](#)]
51. Yang, Y.; Chung, E.K.; Wu, Y.L.; Savelli, S.L.; Nagaraja, H.N.; Zhou, B.; Hebert, M.; Jones, K.N.; Shu, Y.; Kitzmiller, K.; et al. Gene copy-number variation and associated polymorphisms of complement component C4 in human systemic lupus erythematosus (SLE): Low copy number is a risk factor for and high copy number is a protective factor against sle susceptibility in European Americans. *Am. J. Hum. Genet.* **2007**, *80*, 1037–1054. [[PubMed](#)]
52. Levy, S.; Sutton, G.; Ng, P.C.; Feuk, L.; Halpern, A.L.; Walenz, B.P.; Axelrod, N.; Huang, J.; Kirkness, E.F.; Denisov, G.; et al. The diploid genome sequence of an individual human. *PLoS Biol.* **2007**, *5*. [[CrossRef](#)] [[PubMed](#)]
53. Wheeler, D.A.; Srinivasan, M.; Egholm, M.; Shen, Y.; Chen, L.; McGuire, A.; He, W.; Chen, Y.J.; Makhijani, V.; Roth, G.T.; et al. The complete genome of an individual by massively parallel DNA sequencing. *Nature* **2008**, *452*, 872–876. [[CrossRef](#)] [[PubMed](#)]
54. Wellcome Trust Case Control Consortium; Craddock, N.; Hurles, M.E.; Cardin, N.; Pearson, R.D.; Plagnol, V.; Robson, S.; Vukcevic, D.; Barnes, C.; Conrad, D.F.; et al. Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3000 shared controls. *Nature* **2010**, *464*, 713–720. [[PubMed](#)]
55. Zanda, M.; Onengut-Gumuscu, S.; Walker, N.; Shtir, C.; Gallo, D.; Wallace, C.; Smyth, D.; Todd, J.A.; Hurles, M.E.; Plagnol, V.; et al. A genome-wide assessment of the role of untagged copy number variants in type 1 diabetes. *PLoS Genet.* **2014**, *10*. [[CrossRef](#)] [[PubMed](#)]
56. Cooper, N.J.; Shtir, C.J.; Smyth, D.J.; Guo, H.; Swafford, A.D.; Zanda, M.; Hurles, M.E.; Walker, N.M.; Plagnol, V.; Cooper, J.D.; et al. Detection and correction of artefacts in estimation of rare copy number variants and analysis of rare deletions in type 1 diabetes. *Hum. Mol. Genet.* **2015**, *24*, 1774–1790. [[CrossRef](#)] [[PubMed](#)]
57. Hoffman, H.M.; Broderick, L. Editorial: It just takes one: Somatic mosaicism in autoinflammatory disease. *Arthritis Rheumatol.* **2017**, *69*, 253–256. [[CrossRef](#)] [[PubMed](#)]
58. Bruder, C.E.; Piotrowski, A.; Gijsbers, A.A.; Andersson, R.; Erickson, S.; de Stahl, T.D.; Menzel, U.; Sandgren, J.; von Tell, D.; Poplawski, A.; et al. Phenotypically concordant and discordant monozygotic twins display different DNA copy-number-variation profiles. *Am. J. Hum. Genet.* **2008**, *82*, 763–771. [[CrossRef](#)] [[PubMed](#)]
59. Selmi, C.; Cavaciocchi, F.; Lleo, A.; Cheroni, C.; de Francesco, R.; Lombardi, S.A.; de Santis, M.; Meda, F.; Raimondo, M.G.; Crotti, C.; et al. Genome-wide analysis of DNA methylation, copy number variation, and gene expression in monozygotic twins discordant for primary biliary cirrhosis. *Front. Immunol.* **2014**, *5*. [[CrossRef](#)] [[PubMed](#)]
60. Rao, V.K.; Oliveira, J.B. How I treat autoimmune lymphoproliferative syndrome. *Blood* **2011**, *118*, 5741–5751. [[CrossRef](#)] [[PubMed](#)]
61. Holzelova, E.; Vonarbourg, C.; Stolzenberg, M.C.; Arkwright, P.D.; Selz, F.; Prieur, A.M.; Blanche, S.; Bartunkova, J.; Vilmer, E.; Fischer, A.; et al. Autoimmune lymphoproliferative syndrome with somatic Fas mutations. *N. Engl. J. Med.* **2004**, *351*, 1409–1418. [[CrossRef](#)] [[PubMed](#)]
62. Moritake, H.; Takagi, M.; Kinoshita, M.; Ohara, O.; Yamamoto, S.; Moriguchi, S.; Nunoi, H. Autoimmunity including intestinal Behçet disease bearing the KRAS mutation in lymphocytes: A case report. *Pediatrics* **2016**, *137*. [[CrossRef](#)] [[PubMed](#)]



63. Mayhew, A.J.; Meyre, D. Assessing the heritability of complex traits in humans: Methodological challenges and opportunities. *Curr. Genom.* **2017**, *18*, 332–340. [[CrossRef](#)] [[PubMed](#)]
64. Trerotola, M.; Relli, V.; Simeone, P.; Alberti, S. Epigenetic inheritance and the missing heritability. *Hum. Genom.* **2015**, *9*. [[CrossRef](#)] [[PubMed](#)]
65. Wallace, C.; Cutler, A.J.; Pontikos, N.; Pekalski, M.L.; Burren, O.S.; Cooper, J.D.; Garcia, A.R.; Ferreira, R.C.; Guo, H.; Walker, N.M.; et al. Dissection of a complex disease susceptibility region using a bayesian stochastic search approach to fine mapping. *PLoS Genet.* **2015**, *11*. [[CrossRef](#)] [[PubMed](#)]
66. International MHC and Autoimmunity Genetics Network; Rioux, J.D.; Goyette, P.; Vyse, T.J.; Hammarstrom, L.; Fernando, M.M.; Green, T.; de Jager, P.L.; Foisy, S.; et al. Mapping of multiple susceptibility variants within the MHC region for 7 immune-mediated diseases. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 18680–18685. [[PubMed](#)]
67. International Multiple Sclerosis Genetics Consortium; Wellcome Trust Case Control Consortium 2; Sawcer, S.; Hellenthal, G.; Pirinen, M.; Spencer, C.C.; Patsopoulos, N.A.; Moutsianas, L.; Dilthey, A.; Su, Z.; et al. Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature* **2011**, *476*, 214–219. [[CrossRef](#)] [[PubMed](#)]
68. Boyman, O.; Sprent, J. The role of interleukin-2 during homeostasis and activation of the immune system. *Nat. Rev. Immunol.* **2012**, *12*, 180–190. [[CrossRef](#)] [[PubMed](#)]
69. Vella, A.; Cooper, J.D.; Lowe, C.E.; Walker, N.; Nutland, S.; Widmer, B.; Jones, R.; Ring, S.M.; McArdle, W.; Pembrey, M.E.; et al. Localization of a type 1 diabetes locus in the *IL2RA*/*CD25* region by use of tag single-nucleotide polymorphisms. *Am. J. Hum. Genet.* **2005**, *76*, 773–779. [[CrossRef](#)] [[PubMed](#)]
70. Lowe, C.E.; Cooper, J.D.; Brusko, T.; Walker, N.M.; Smyth, D.J.; Bailey, R.; Bourget, K.; Plagnol, V.; Field, S.; Atkinson, M.; et al. Large-scale genetic fine mapping and genotype-phenotype associations implicate polymorphism in the *IL2RA* region in type 1 diabetes. *Nat. Genet.* **2007**, *39*, 1074–1082. [[CrossRef](#)] [[PubMed](#)]
71. Maier, L.M.; Lowe, C.E.; Cooper, J.; Downes, K.; Anderson, D.E.; Severson, C.; Clark, P.M.; Healy, B.; Walker, N.; Aubin, C.; et al. *IL2RA* genetic heterogeneity in multiple sclerosis and type 1 diabetes susceptibility and soluble interleukin-2 receptor production. *PLoS Genet.* **2009**, *5*. [[CrossRef](#)] [[PubMed](#)]
72. Dendrou, C.A.; Plagnol, V.; Fung, E.; Yang, J.H.; Downes, K.; Cooper, J.D.; Nutland, S.; Coleman, G.; Himsforth, M.; Hardy, M.; et al. Cell-specific protein phenotypes for the autoimmune locus *IL2RA* using a genotype-selectable human bioresource. *Nat. Genet.* **2009**, *41*, 1011–1015. [[CrossRef](#)] [[PubMed](#)]
73. Oksenberg, J.R.; Barcellos, L.F.; Cree, B.A.; Baranzini, S.E.; Bugawan, T.L.; Khan, O.; Lincoln, R.R.; Swerdlin, A.; Mignot, E.; Lin, L.; et al. Mapping multiple sclerosis susceptibility to the *HLA-DR* locus in African Americans. *Am. J. Hum. Genet.* **2004**, *74*, 160–167. [[CrossRef](#)] [[PubMed](#)]
74. Noble, J.A.; Martin, A.; Valdes, A.M.; Lane, J.A.; Galgani, A.; Petrone, A.; Lorini, R.; Pozzilli, P.; Buzzetti, R.; Erlich, H.A. Type 1 diabetes risk for human leukocyte antigen (HLA)-DR3 haplotypes depends on genotypic context: Association of DPB1 and HLA class I loci among DR3- and DR4-matched Italian patients and controls. *Hum. Immunol.* **2008**, *69*, 291–300. [[CrossRef](#)] [[PubMed](#)]
75. Gough, S.C.; Simmonds, M.J. The HLA region and autoimmune disease: Associations and mechanisms of action. *Curr. Genom.* **2007**, *8*, 453–465.
76. Nakanishi, K.; Inoko, H. Combination of HLA-A24, -DQA1\*03, and -DR9 contributes to acute-onset and early complete  $\beta$ -cell destruction in type 1 diabetes: Longitudinal study of residual beta-cell function. *Diabetes* **2006**, *55*, 1862–1868. [[CrossRef](#)] [[PubMed](#)]
77. Lipponen, K.; Gombos, Z.; Kiviniemi, M.; Siljander, H.; Lempainen, J.; Hermann, R.; Veijola, R.; Simell, O.; Knip, M.; Ilonen, J. Effect of HLA class I and class II alleles on progression from autoantibody positivity to overt type 1 diabetes in children with risk-associated class II genotypes. *Diabetes* **2010**, *59*, 3253–3256. [[CrossRef](#)] [[PubMed](#)]
78. Hughes, T.; Adler, A.; Kelly, J.A.; Kaufman, K.M.; Williams, A.H.; Langefeld, C.D.; Brown, E.E.; Alarcon, G.S.; Kimberly, R.P.; Edberg, J.C.; et al. Evidence for gene-gene epistatic interactions among susceptibility loci for systemic lupus erythematosus. *Arthritis Rheum.* **2012**, *64*, 485–492. [[CrossRef](#)] [[PubMed](#)]
79. Galarza-Munoz, G.; Briggs, F.B.S.; Evsyukova, I.; Schott-Lerner, G.; Kennedy, E.M.; Nyanhete, T.; Wang, L.; Bergamaschi, L.; Widen, S.G.; Tomaras, G.D.; et al. Human epistatic interaction controls *IL7R* splicing and increases multiple sclerosis risk. *Cell* **2017**, *169*, 72–84. [[CrossRef](#)] [[PubMed](#)]
80. Visscher, P.M.; Wray, N.R.; Zhang, Q.; Sklar, P.; McCarthy, M.I.; Brown, M.A.; Yang, J. 10 years of GWAS discovery: Biology, function, and translation. *Am. J. Hum. Genet.* **2017**, *101*, 5–22. [[CrossRef](#)] [[PubMed](#)]

81. Gauderman, W.J. Sample size requirements for association studies of gene-gene interaction. *Am. J. Epidemiol.* **2002**, *155*, 478–484. [[CrossRef](#)] [[PubMed](#)]
82. Gaunt, T.R.; Shihab, H.A.; Hemani, G.; Min, J.L.; Woodward, G.; Lyttleton, O.; Zheng, J.; Duggirala, A.; McArdle, W.L.; Ho, K.; et al. Systematic identification of genetic influences on methylation across the human life course. *Genom. Biol.* **2016**, *17*. [[CrossRef](#)] [[PubMed](#)]
83. Richardson, T.G.; Shihab, H.A.; Hemani, G.; Zheng, J.; Hannon, E.; Mill, J.; Carnero-Montoro, E.; Bell, J.T.; Lyttleton, O.; McArdle, W.L.; et al. Collapsed methylation quantitative trait loci analysis for low frequency and rare variants. *Hum. Mol. Genet.* **2016**, *25*, 4339–4349. [[CrossRef](#)] [[PubMed](#)]
84. Imgenberg-Kreuz, J.; Sandling, J.K.; Almlöf, J.C.; Nordlund, J.; Signer, L.; Norheim, K.B.; Omdal, R.; Ronnblom, L.; Eloranta, M.L.; Syvanen, A.C.; et al. Genome-wide DNA methylation analysis in multiple tissues in primary Sjögren's syndrome reveals regulatory effects at interferon-induced genes. *Ann. Rheum. Dis.* **2016**, *75*, 2029–2036. [[CrossRef](#)] [[PubMed](#)]
85. Lemire, M.; Zaidi, S.H.; Ban, M.; Ge, B.; Aissi, D.; Germain, M.; Kassam, I.; Wang, M.; Zanke, B.W.; Gagnon, F.; et al. Long-range epigenetic regulation is conferred by genetic variation located at thousands of independent loci. *Nat. Commun.* **2015**, *6*. [[CrossRef](#)] [[PubMed](#)]
86. Belot, M.P.; Fradin, D.; Mai, N.; Le Fur, S.; Zelenika, D.; Kerr-Conte, J.; Pattou, F.; Lucas, B.; Bougneres, P. CpG methylation changes within the *IL2RA* promoter in type 1 diabetes of childhood onset. *PLoS ONE* **2013**, *8*. [[CrossRef](#)] [[PubMed](#)]
87. Richardson, T.G.; Zheng, J.; Smith, G.D.; Timpson, N.J.; Gaunt, T.R.; Relton, C.L.; Hemani, G. Mendelian randomization analysis identifies CpG sites as putative mediators for genetic influences on cardiovascular disease risk. *Am. J. Hum. Genet.* **2017**, *101*, 590–602. [[CrossRef](#)] [[PubMed](#)]
88. Millstein, J.; Zhang, B.; Zhu, J.; Schadt, E.E. Disentangling molecular relationships with a causal inference test. *BMC Genet.* **2009**, *10*. [[CrossRef](#)] [[PubMed](#)]
89. Relton, C.L.; Smith, G.D. Two-step epigenetic mendelian randomization: A strategy for establishing the causal role of epigenetic processes in pathways to disease. *Int. J. Epidemiol.* **2012**, *41*, 161–176. [[CrossRef](#)] [[PubMed](#)]
90. Relton, C.L.; Smith, G.D. Mendelian randomization: Applications and limitations in epigenetic studies. *Epigenomics* **2015**, *7*, 1239–1243. [[CrossRef](#)] [[PubMed](#)]
91. Ladd-Acosta, C.; Fallin, M.D. The role of epigenetics in genetic and environmental epidemiology. *Epigenomics* **2016**, *8*, 271–283. [[CrossRef](#)] [[PubMed](#)]
92. Liu, Y.; Aryee, M.J.; Padyukov, L.; Fallin, M.D.; Hesselberg, E.; Runarsson, A.; Reinius, L.; Acevedo, N.; Taub, M.; Ronninger, M.; et al. Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis. *Nat. Biotechnol.* **2013**, *31*, 142–147. [[CrossRef](#)] [[PubMed](#)]
93. Olsson, A.H.; Volkov, P.; Bacos, K.; Dayeh, T.; Hall, E.; Nilsson, E.A.; Ladenvall, C.; Ronn, T.; Ling, C. Genome-wide associations between genetic and epigenetic variation influence mRNA expression and insulin secretion in human pancreatic islets. *PLoS Genet.* **2014**, *10*. [[CrossRef](#)] [[PubMed](#)]
94. Jody, Y.; Richardson, T.G.; McArdle, W.; Relton, C.; Gillespie, K.M.; Suderman, M.; Hemani, G. Identification of loci where DNA methylation potentially mediate genetic risk of type 1 diabetes. *J. Autoimmunity* **2018**. [[CrossRef](#)]
95. Richardson, T.G.; Haycock, P.C.; Zheng, J.; Timpson, N.J.; Gaunt, T.R.; Smith, G.D.; Relton, C.L.; Hemani, G. Systematic mendelian randomization framework elucidates hundreds of CpG sites which may mediate the influence of genetic variants on disease. *Hum. Mol. Genet.* **2018**. [[CrossRef](#)] [[PubMed](#)]
96. Richardson, T.G.; Haycock, P.C.; Zheng, J.; Timpson, N.J.; Gaunt, T.R.; Smith, G.D.; Relton, C.L.; Hemani, G. Systematic mendelian randomization framework elucidates hundreds of genetic loci which may influence disease through changes in DNA methylation levels. *bioRxiv* **2017**. [[CrossRef](#)]
97. Cucca, F.; Goy, J.V.; Kawaguchi, Y.; Esposito, L.; Merriman, M.E.; Wilson, A.J.; Cordell, H.J.; Bain, S.C.; Todd, J.A. A male-female bias in type 1 diabetes and linkage to chromosome Xp in MHC HLA-DR3-positive patients. *Nat. Genet.* **1998**, *19*, 301–302. [[CrossRef](#)] [[PubMed](#)]
98. Chabchoub, G.; Uz, E.; Maalej, A.; Mustafa, C.A.; Rebai, A.; Mnif, M.; Bahloul, Z.; Farid, N.R.; Ozcelik, T.; Ayadi, H. Analysis of skewed X-chromosome inactivation in females with rheumatoid arthritis and autoimmune thyroid diseases. *Arthritis Res. Ther.* **2009**, *11*. [[CrossRef](#)] [[PubMed](#)]

99. Simmonds, M.J.; Kavvoura, F.K.; Brand, O.J.; Newby, P.R.; Jackson, L.E.; Hargreaves, C.E.; Franklyn, J.A.; Gough, S.C. Skewed X chromosome inactivation and female preponderance in autoimmune thyroid disease: An association study and meta-analysis. *J. Clin. Endocrinol. Metab.* **2014**, *99*, E127–E131. [[CrossRef](#)] [[PubMed](#)]
100. Milligan, M.J.; Lipovich, L. Pseudogene-derived lncRNAs: Emerging regulators of gene expression. *Front. Genet.* **2014**, *5*. [[CrossRef](#)] [[PubMed](#)]
101. Kung, J.T.; Colognori, D.; Lee, J.T. Long noncoding RNAs: Past, present, and future. *Genetics* **2013**, *193*, 651–669. [[CrossRef](#)] [[PubMed](#)]
102. Liu, W.; Ding, C. Roles of lncRNAs in viral infections. *Front. Cell. Infect. Microbiol.* **2017**, *7*. [[CrossRef](#)] [[PubMed](#)]
103. Spurlock, C.F.; Tossberg, J.T., 3rd; Guo, Y.; Collier, S.P.; Crooke, P.S., 3rd; Aune, T.M. Expression and functions of long noncoding RNAs during human T helper cell differentiation. *Nat. Commun.* **2015**, *6*. [[CrossRef](#)] [[PubMed](#)]
104. Pearson, M.J.; Jones, S.W. Review: Long noncoding RNAs in the regulation of inflammatory pathways in rheumatoid arthritis and osteoarthritis. *Arthritis Rheumatol.* **2016**, *68*, 2575–2783. [[CrossRef](#)] [[PubMed](#)]
105. Mirza, A.H.; Kaur, S.; Brorsson, C.A.; Pociot, F. Effects of GWAS-associated genetic variants on lncRNAs within IBD and T1D candidate loci. *PLoS ONE* **2014**, *9*. [[CrossRef](#)] [[PubMed](#)]
106. Kumar, V.; Westra, H.J.; Karjalainen, J.; Zhernakova, D.V.; Esko, T.; Hrdlickova, B.; Almeida, R.; Zhernakova, A.; Reinmaa, E.; Vosa, U.; et al. Human disease-associated genetic variation impacts large intergenic non-coding RNA expression. *PLoS Genet.* **2013**, *9*. [[CrossRef](#)] [[PubMed](#)]
107. Hrdlickova, B.; Kumar, V.; Kanduri, K.; Zhernakova, D.V.; Tripathi, S.; Karjalainen, J.; Lund, R.J.; Li, Y.; Ullah, U.; Modderman, R.; et al. Expression profiles of long non-coding RNAs located in autoimmune disease-associated regions reveal immune cell-type specificity. *Genom. Med.* **2014**, *6*. [[CrossRef](#)] [[PubMed](#)]
108. Pennacchio, L.A.; Bickmore, W.; Dean, A.; Nobrega, M.A.; Bejerano, G. Enhancers: Five essential questions. *Nat. Rev. Genet.* **2013**, *14*, 288–295. [[CrossRef](#)] [[PubMed](#)]
109. Farh, K.K.; Marson, A.; Zhu, J.; Kleinewietfeld, M.; Housley, W.J.; Beik, S.; Shores, N.; Whitton, H.; Ryan, R.J.; Shishkin, A.A.; et al. Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* **2015**, *518*, 337–343. [[CrossRef](#)] [[PubMed](#)]
110. Carella, C.; Mazziotti, G.; Amato, G.; Braverman, L.E.; Roti, E. Clinical review 169: Interferon- $\alpha$ -related thyroid disease: Pathophysiological, epidemiological, and clinical aspects. *J. Clin. Endocrinol. Metab.* **2004**, *89*, 3656–3661. [[CrossRef](#)] [[PubMed](#)]
111. Cabrera, S.M.; Chen, Y.G.; Hagopian, W.A.; Hessner, M.J. Blood-based signatures in type 1 diabetes. *Diabetologia* **2016**, *59*, 414–425. [[CrossRef](#)] [[PubMed](#)]
112. Psarras, A.; Emery, P.; Vital, E.M. Type I interferon-mediated autoimmune diseases: Pathogenesis, diagnosis and targeted therapy. *Rheumatology* **2017**, *56*, 1662–1675. [[CrossRef](#)] [[PubMed](#)]
113. Stefan, M.; Wei, C.; Lombardi, A.; Li, C.W.; Concepcion, E.S.; Inabnet, W.B., 3rd; Owen, R.; Zhang, W.; Tomer, Y. Genetic-epigenetic dysregulation of thymic TSH receptor gene expression triggers thyroid autoimmunity. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 12562–12567. [[CrossRef](#)] [[PubMed](#)]
114. Lin, J.D.; Cheng, C.W. Causal variants in autoimmune disease: A commentary on a recent published fine-mapping algorithm analysis in genome-wide association studies study. *Ann. Transl. Med.* **2017**, *5*. [[CrossRef](#)] [[PubMed](#)]
115. Perez-Pinera, P.; Kocak, D.D.; Vockley, C.M.; Adler, A.F.; Kabadi, A.M.; Polstein, L.R.; Thakore, P.I.; Glass, K.A.; Ousterout, D.G.; Leong, K.W.; et al. RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nat. Methods* **2013**, *10*, 973–976. [[CrossRef](#)] [[PubMed](#)]
116. Simeonov, D.R.; Gowen, B.G.; Boontanart, M.; Roth, T.L.; Gagnon, J.D.; Mumbach, M.R.; Satpathy, A.T.; Lee, Y.; Bray, N.L.; Chan, A.Y.; et al. Discovery of stimulation-responsive immune enhancers with CRISPR activation. *Nature* **2017**, *549*, 111–115. [[CrossRef](#)] [[PubMed](#)]
117. Warwick, A.; Mackay, I.R. Intolerant bodies: A short history of autoimmunity. In *Johns Hopkins Biographies of Disease*; Johns Hopkins University Press: Baltimore, MD, USA, 2014.

